Estimation of vessel collision risk index based on support vector machine

Longhui Gang[1], Yonghui Wang[2], Yao Sun[2], Liping Zhou[3] and Mingheng Zhang[4]

Abstract

Collision risk index is important for assessing vessel collision risk and is one of the key problems in the research field of vessel collision avoidance. With accurate collision risk index obtained through vessel movement parameters and encounter situation analysis, the pilot can adopt correct avoidance action. In this article, a collision risk index estimation model based on support vector machine is proposed. The proposed method comprises two units, that is, support vector machine–based unit for predicting the collision risk index and the genetic algorithm–based unit for optimizing the parameters of support vector machine. The model and algorithm are illustrated in the empirical analysis phase, and the comparison results show that genetic algorithm-support vector machine model can generally provide a better performance for collision risk index estimation. Meanwhile, the result also indicates that the model may be not so good when we take a higher value of collision risk index. So, the distinguishing threshold of collision risk level should be adjusted according to actual situation when applying this model in practical application.

Keywords

Collision avoidance, collision risk index, support vector machine, genetic algorithm

Introduction

Background

Collision, grounding, and striking on rocks are the common accidents when vessels are navigating. Once accident happened, the serious casualties, property loss, and environmental pollution are inevitable. Especially for the vessels such as oil tanker, liquefied petroleum gas (LPG), liquefied natural gas (LNG), chemical tanker, and nuclear ship, vessels collision will lead to serious marine environmental pollution and irreparable ecological disaster. For example, on 8 March 1967, oil tanker Torry Canyon ran aground near the southeast coastal areas of Britain. This accident led to crude oil leak of 100 thousand tons and the clean-up costs reached up to 10 million ponds. On 6 March 1978, super oil tanker Amoco Cadiz ran aground near the French Brittany seas, causing 220 thousand tons of crude oil leaking and polluting 180 km French seacoast. The clean-up costs reached up to 100 million dollars, fishery damages to 3000 thousand, and travel damages to 60 million dollars. So, vessel accident, especially dangerous cargo vessel accident, not only results in casualties and property loss but also brings marine environment and marine organism with serious

[1]School of Navigation, Dalian Maritime University, Dalian, China
[2]Transportation Management College, Dalian Maritime University, Dalian, China
[3]Wuxi Mingda Traffic & Technology Consulted Co., Ltd., Wuxi, China
[4]School of Automotive Engineering, Dalian University of Technology, Dalian, China

Corresponding author:
Mingheng Zhang, School of Automotive Engineering, Dalian University of Technology, Dalian 116024, China.
Email: gloriazhang@163.com

ecological disasters. At present, how to avoid and reduce the damage caused by vessel accident has been becoming an extremely significant problem.

In navigation safety research field, vessel collision avoidance is an important problem to be solved. In which, as the basis index of assessing vessel encounter situation, collision risk index (CRI) is essential for the pilot adopting correct avoidance action and for the engineers developing intelligence vessel collision avoidance system. In fact, CRI is influenced by many factors, such as distance of close point of approaching (DCPA), time to close point of approaching (TCPA), azimuth of coming vessels, vessels velocity, and the situation of visibility distance. Therefore, how to calculate the CRI index according to the vessels movement and encounter situation is a challenging task.

Literature review

In the past decades, CRI has been studied under the Convention on the International Regulations for Preventing Collisions at Sea (COLREGs) since 1972. The main methods include DCPA and TCPA weighting measure method,[1–7] fuzzy comprehensive evaluation method,[8–13] artificial neural network (ANN) method,[14–19] and other new methods.[20–25]

At the beginning of the research on the vessel collision risk, based on initial DCPA and TCPA data, researchers focus on the factor's weight determination and their application with distributed regularity. For instance, Kearon[1] proposed a CRI estimation method through parameters weighting and then the collision risk was determined with practical experiences of collision at sea.

However, from the perspective of practice, the process of collision risk judgment has the features of fuzzy and ambiguity. How to evaluate CRI based on actual pilot decision process is an important problem need to be resolved. Therefore, AN Cockeroft[8] applied fuzzy reasoning theory to the presented collision avoidance model. At that time, this method is an advanced technique, but there are still certain difficulties in determining the membership function. Based on fuzzy comprehensive evaluation method, Kim and Kim[12] presented a CRI model considering the factors of DCPA, TCPA, the target ship distance, locations, speed ratio, and touch angle. But, the factors such as human, vessels, and environment are not included. On this basis, Zhou and Wu[10] improved this model and introduced the vessel type and tonnage, pilot skill and visibility, and other factors into this research field. In consideration of the nonlinear and complexity characteristic for CRI estimation, Chen[14] proposed a CRI estimation model combining ANN with fuzzy reasoning theory. This method has the traits that fuzzy logic can be understood easily and ANN has extremely strong adaptive capacity.

With the development of research on vessel collision risk avoidance, accurate CRI level classification plays more and more important roles for the effective ship collision avoidance strategy making. Chin and Debnath[20] proposed a binomial model based on DCPA and TCPA, which classifies vessels' encounter situation into serious and low-risk situation. And then CRI was calculated by ordered probability unit regression model, in which collision risk is considered as continuous variable and effected by DCPA and TCPA monotonously. Ren et al.[5] raised a linear model which considered several variables such as vessel type, velocity and route as the basic input vector for evaluating collision risk. Furthermore, on the basis of this method, considering TCPA and relative bearing between two vessels, an evaluation method based on multiplication of basic collision risk, TCPA, DCPA, and angle is proposed in this article.

In summary, the scholars have profoundly studied in CRI. The research about employing fuzzy comprehensive evaluation to CRI determination is more than others. Fuzzy comprehensive evaluation is used to calculate relative movement parameters between vessels according to initial data of vessel (such as DCPA and TCPA) and then the parameters are used as the evaluation index for CRI calculation through fuzzy inference rules or fuzzy membership functions. However, this method still has some certain limitations when the calculation is complicated. Since there are many factors affecting CRI determination (such as velocity and direction), relationship between the factors and CRI has the characteristics such as nonlinearity, stochasticity, and ambiguity. So, how to solve the problem from multiple attribute decision making process is a big challenge task.

Support vector machine (SVM) is a relatively new kind of learning machine which has been applied successfully to estimation.[26] Considering the parameters will greatly affect the performance of SVM, some researchers attempted to determine the parameter values for their problems. Hou and Li[27] attempted to determine the parameters in SVM by use of evolution strategy with covariance matrix adaptation. Hsu et al.[28] applied grid search to determine the adaptive values for the parameters in SVM. Heuristic algorithms have been successfully used in many complex problems.[29–31] Thus, this article also applies a common heuristic algorithm,[32] genetic algorithm (GA) to determine the appropriate parameters for SVM.

The purpose of this article is building a CRI estimation model based on SVM, which is used to solve the complicated nonlinear mapping problem between CRI and dynamical vessel parameters. The article is organized as follows: in section "CRI estimation based
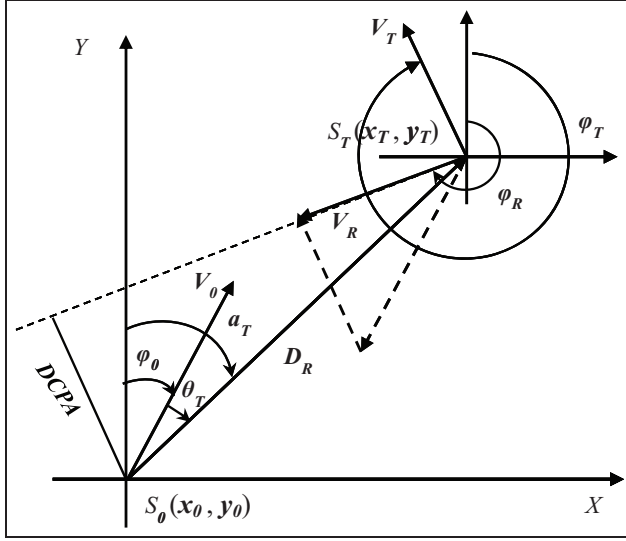
Figure 1. The diagram of vessel collision geometry.

on fuzzy comprehensive evaluation," the impacting factors description and calculation based on fuzzy comprehensive evaluation are demonstrated. In section "Estimation model of CRI based on SVM," CRI estimation model based on SVM is proposed and the parameters are optimized with GA method. In section "Simulation analysis," the presented model is tested and results are discussed. Finally, the conclusions are given together with suggestions for further study.

CRI estimation based on fuzzy comprehensive evaluation

Influence factors analysis

In the process of vessel collision avoidance, CRI can be used to measure the danger degree of collision. DCPA and TCPA are usually considered as the most effective index for determining the value of CRI. As shown in Figure 1, DCPA and TCPA can be obtained by geometric calculation of vessel collision avoidance.

Given the own ship coordinate, direction, and velocity as $S_0(x_0, y_0)$, ϕ_0 and V_0, the target ship as $S_T(x_T, y_T)$, ϕ_T, and V_T. The relative moving parameters can be obtained as

$$D_R = \sqrt{(X_T - X_0)^2 + (Y_T - Y_0)^2} \quad (1)$$

$$V_R = V_0 \times \sqrt{1 + \left(\frac{V_T}{V_0}\right)^2 - 2 \cdot \frac{V_T}{V_0} \cdot \cos(\phi_0 - \phi_T)} \quad (2)$$

$$\phi_R = \cos^{-1}\left(\frac{V_0 - V_T \cdot \cos(\phi_0 - \phi_T)}{V_R}\right) \quad (3)$$

$$DCPA = D_R \times \sin(\phi_R - \alpha_T - \pi) \quad (4)$$

$$TCPA = D_R \times \cos(\phi_R - \alpha_T - \pi)/V_R \quad (5)$$

where D_R denotes the relative distance between the owner and target ships, V_R denotes the relative velocity, ϕ_R denotes the relative course, α_T denotes the azimuth of the target ship, and θ_T denotes the relative bearing.

When $DCPA$ equals 0, it means that if the two vessels maintain the current velocity and direction, the collision will happen after a certain period at one point. $DCPA > 0$ indicates that there is a certain distance between two vessels when both of them are encountering. But there may still exist collision risk. If $TCPA$ takes a larger value, the danger degree of collision cannot be induced only by $DCPA$. If $DCPA$ equals 0 or less than a given value, the smaller the value of $TCPA$, the larger the degree of collision danger, and vice versa. So, in collision avoidance practice, if DCPA is less than a safe distance and TCPA is smaller, there may be a risk of collision.

DCPA and TCPA are the most main and directive factors for judging the collision existing or not. On the other hand, DCPA and D_R are the most practical factors in the actual practice. In order to obtain the degree of collision danger, only considering DCPA and TCPA or DCPA and D_R is not enough. We should simultaneously consider the effect of other factors, such as relative orientation of two vessels and relative bearing.

CRI calculation

CRI has the characteristic of ambiguity and complexity. So it can be determined with fuzzy comprehensive evaluation theory. It can be represented as

$$CRI = W \cdot U = (w_{DCPA}, w_{TCPA}, w_{D_R}, w_{\theta_T}, w_K) \begin{bmatrix} u_{DCPA} \\ u_{TCPA} \\ u_{D_R} \\ u_{\theta_T} \\ u_K \end{bmatrix} \quad (6)$$

where U is the membership matrix of target factor, W is weight matrix, and K is velocity ratio of two vessels, $K = V_T/V_0$.
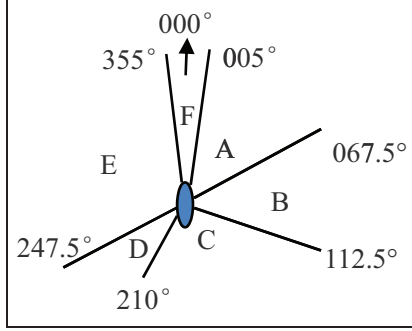
1. Membership function of DCPA

The larger the value of DCPA, the smaller the degree of collision danger. The membership function of DCPA can be presented as

$$u_{DCPA} = \begin{cases} 1 & , |DCPA| < d_1 \\ \left(\frac{d_2 - |DCPA|}{d_2 - d_1}\right)^2 & , d_1 \leq |DCPA| \leq d_2 \\ 0 & , |DCPA| > d_2 \end{cases} \quad (7)$$

Table 1. Observation value of d_1 and d_2.

Relative bearing θ_T	355°–067.5°	067.5°–112.5°	112.5°–247.5°	247.5°–355°
d_1 (nautical miles)	1.1	1.0	0.6	0.9
d_2 (nautical miles)	2.2	2.0	1.2	1.8

Figure 2. Situation of vessel collision avoidance.

Figure 3. Distribution figure of θ_T.

where d_1 is the minimal safe encounter distance and d_2 is the absolute safe encounter distance. The range of d_1 and d_2 can refer to the amended Goodwin observation data, as shown in Figures 2 and 3 and Table 1.

2. Risk membership function of TCPA

The risk membership function of TCPA can be represented as

$$u_{TCPA} = \begin{cases} 1 & ,0 \le |TCPA| \le t_1 \\ \left(\dfrac{t_2 - |TCPA|}{t_2 - t_1}\right)^2 & ,t_1 < |TCPA| \le t_2 \\ 0 & ,|TCPA| > t_2 \end{cases} \quad (8)$$
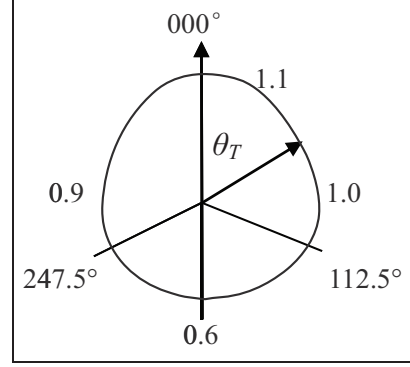
where the range of t_1 and t_2 can be represented as

$$t_1 = \begin{cases} \dfrac{\sqrt{d_1^2 - DCPA^2}}{V_R} & ,DCPA \le d_1 \\ \dfrac{d_1 - DCPA}{V_R} & ,DCPA > d_1 \end{cases}, \quad t_2 = \dfrac{\sqrt{d_2^2 - DCPA^2}}{V_R} \quad (9)$$

3. Risk membership function of D_R

The smaller the D_R, the smaller the distance between the target vessel and the own vessel, and the greater the degree of collision risk. The risk membership function of D_R can be presented as

$$u_{D_R} = \begin{cases} 1 & ,0 < D_R < D_1 \\ \left(\dfrac{D_2 - D_R}{D_2 - D_1}\right)^2 & ,D_1 \le D_R \le D_2 \\ 0 & ,D_R > D_2 \end{cases} \quad (10)$$

where D_1 is the critical safe distance, which is usually equal to 12 times length of vessel. D_2 is the distance in which pilots can adapt avoidance measure. The value of D_2 is usually equal to R which is the radius of marine power model obtained by Davis

$$R = 1.7\cos(\theta_T - 19°) + \sqrt{4.4 + 2.89\cos^2(\theta_T - 19°)} \quad (11)$$

4. Membership function of θ_T

For the owner ship, target vessels at different position have different effect on the risk level of collision. Generally, the danger from right side is larger than the left side and the front is larger than the back. The membership function of θ_T can be represented as

$$u_{\theta_T} = \frac{1}{2}\left[\cos(\theta_T - 19°) + \sqrt{\frac{440}{289} + \cos^2(\theta_T - 19°)}\right] - \frac{5}{17} \quad (12)$$

5. Membership function of K

The membership function of K can be represented as

$$u_K = \frac{1}{1 + \dfrac{2}{K\sqrt{K^2 + 1 + 2K\sin C}}} \quad (13)$$

Estimation model of CRI based on SVM

In the actual practice of navigation, the vessel's initial moving parameters and partial relative parameters can be obtained by onboard instrument. The pilots should identify the existing risk through fusing the information presented with these parameters. However, the relation between CRI and vessel's moving parameters is complicated and nonlinear. Therefore, how to describe these characters and propose a corresponding model for CRI level classification is a big problem need to be solved.

As a new and promising technique for classification and regression problems, SVM can be adjusted to map the complex input–output relationship for the nonlinear system without dependent on the specific functions. Therefore, considering the actual needs in this research and the merit of SVM, this article applies SVM to the estimation of CRI.

Support vector regression

Given $x_i \in R^n$, $y_i \in R$ as the input and output vector for SVM. Then, x is mapped to high-dimension feature space through nonlinear mapping function Φ

$$f(x) = \omega\phi(x) + b \qquad (14)$$

The optimized regression function is the minimum and regularized generic function under certain constraints

$$\frac{1}{2}\|\omega\|^2 + C\frac{1}{l}\sum_{i=1}^{l}L_\epsilon(y_i, f(x_i)) \qquad (15)$$

where $\|\omega\|$ is named as the regularized term, which makes the function flat and improve its generalization ability; $C\frac{1}{l}\sum_{i=1}^{l}(\xi_i + \xi_i^*)$ $(C>0)$ is named as the experience risk generic function, which can be determined by different loss functions; C is used to balance the relationship between structure risks and experience risks. Due to the request of balancing structural risk and empirical risk, non-negative slack variables ξ_i, ξ_i^* are introduced describing this balancing. Formula (15) can be represented with

$$\min\frac{1}{2}\|\omega\|^2 + C\frac{1}{l}\sum_{i=1}^{l}(\xi_i + \xi_i^*) \qquad (16)$$

$$\text{s.t.} \, y_i - \omega\phi(x) - b \le \varepsilon + \xi_i \qquad (17)$$

$$\omega\phi(x) + b - y_i \le \varepsilon + \xi_i^* \qquad (18)$$

$$\xi_i^* \ge 0, i = 1, \ldots, 1$$

The minimization of formula (17) is a convex quadratic optimization problem and then the problem can be inferred with the Lagrange multipliers

$$\omega - \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)x_i = 0 \qquad (19)$$

Then

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\phi(x_i) \cdot \phi(x) + b \qquad (20)$$

Substituting the kernel function into formula (20), then

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K(x_i, x) + b \qquad (21)$$

where $K(x_i, x_j)$ is the inner product of vectors x_i and x_j on the feature space $\phi(x_i)$ and $\phi(x_j)$, respectively.

Structure of the SVM model

In the actual practice of navigation, pilots can directly obtain the owner vessel and the target vessel's initial moving parameters such as ϕ_0, ϕ_T, V_0, and V_T through onboard instrument such as Automatic Identification System (AIS). The relative bearing θ_T and distance D_R can be obtained with radar sensor. Among the factors for risk level estimation, the length of vessel (LOA) is known in advance, the time for turning 90°, and shifting distance of gravity and advance can be determined by the owner vessel's situation. The evaluation indexes of CRI, such as DCPA and TCPA, can be obtained by two encountering vessels' original data. Therefore, in this article, the CRI estimation model based on SVM is proposed as Figure 4. In which, each of the parameters above are combined as an input vector $V_{in} = [\varphi_0, \varphi_T, V_0, V_t, \theta_T, D_R]$, and the value of CRI obtained by fuzzy comprehensive evaluation is used as output.

Parameter optimization

Although SVM is feasible for CRI estimation, there are some parameters, which greatly impact the performance of SVM, that need to be optimized in advance. In general, for radial basis function (RBF) kernel,
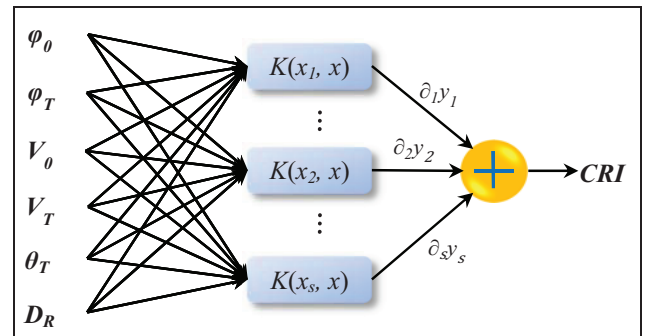

Figure 4. CRI estimation model based on SVM.

parameters C, γ, and ε are very important for the SVM prediction performance. So, the parameter optimization is essential for improving the estimation accuracy. Furthermore, there are several parameter optimization methods such as cross-validation (CV), particle swarm optimization (PSO), and GA. In which, GA is a search heuristic that mimics the process of natural selection and is also tested by lots of researchers as an effective method to solve this kind of complex problems. So, this article proposes a GA-SVM model which applies GA for the parameter optimization. The key steps are as follows:

1. *Chromosome coding.* Assume that the parameters C, γ, and ε in each chromosome can be represented as $gene_1^g$, $gene_2^g$, and $gene_3^g$, in which g is the current generation. In order to reduce the complexity of space search, the range of C, γ, and ε are limited to: $C \in [2^{-5}, 2^5]$, $\varepsilon \in [2^{-13}, 2^{-1}]$, and $\gamma \in [0, 2]$.

2. *Fitness function.* In this article, considering that GA is always finding the maximum fitness of the individual chromosome, mean squared error (MSE) is adopted

$$\text{fitness} = \frac{1}{l} \sum_{i=1}^{l} \left(f(x_i) - y_i \right)^2 \qquad (22)$$

where $f(x_i)$ is the objective value given by SVM, y_i is the observation value, and l is the number of observation variables.

3. *Selection operation.* In order to reduce the computation time, here we choose Roulette selection strategy as the selection operation and implement best part of the chromosome retention strategy (i.e. in each generation the partial individual who has highest fitness directly becomes the next-generation population). The strategies not only guarantee convergence of algorithm but also increase the pressure of selection and accelerate the convergence. The specific operational process is as follows: we set selection parameters P_S and set P_S/N (N is the scale of population) as threshold. The chromosome whose adaptation rate is bigger than the threshold can directly retain to the next generation without Roulette selection. Otherwise, if chromosome can be chosen to the next generation depends on the Roulette selection.

4. *Crossover operation.* The crossover operation produces offspring through exchanging two parent chromosomes' gene. The specific crossover operation is as follows

$$\begin{cases} e_l^{child,1} = \left[\delta_l' e_l^{parent,1} + \left(1 - \delta_l'\right) e_l^{parent,2} \right], \\ e_l^{child,2} = \left[\delta_l' e_l^{parent,2} + \left(1 - \delta_l'\right) e_l^{parent,1} \right], \quad \delta_l'' < p_c \\ \qquad e_l^{child,1} = e_l^{parent,1}, \\ \qquad e_l^{child,2} = e_l^{parent,2}, \qquad\qquad\qquad \text{otherwise} \end{cases}$$

$$(23)$$

where $e_l^{parent,1}$ and $e_l^{parent,2}$ represent lth gene of two parent chromosomes, respectively; $e_l^{child,1}$ and $e_l^{child,2}$ represent lth gene of two offspring chromosomes, respectively; δ_l' and δ_l'' represent random figure between 0 and 1; and p_c is crossover rate.

5. *Mutation operation.* The number of mutation is controlled by the mutation rate p_m. If lth gene of parent chromosome is chosen to mutate, then

$$\begin{cases} e_l^{child} = \left[e_l^{parent} \times \left(1 + \delta_l'''\right)^{(1 - \tau/\tau^{max})\lambda} \right], \delta_l'''' < p_c \\ \\ e_l^{child} = \left[e_l^{parent} \times \left(1 - \delta_l'''\right)^{(1 - \tau/\tau^{max})\lambda} \right], \text{otherwise} \end{cases}$$

$$(24)$$

where e_l^{parent} represents lth gene of parent chromosome; e_l^{child} represents lth gene of offspring chromosome; δ_l''' and δ_l'''' represent random figure between 0 and 1, respectively; p_m is mutation rate; and τ^{max} and τ represent maximum evolution generation and current evolution generation, respectively. In addition, at the initial stage of algorithm, the bigger mutation degree of chromosome means abundant population and that algorithm can search solution in more broad scale. But at the last stage, the bigger mutation degree of chromosome influences the speed of convergence. So we set parameter $\lambda (2 \le \lambda \le 5)$ to control the effect evolution generation exerts on mutation degree.

6. *Termination condition.* In this article, the search loop continues until $MSE_n - MSE_{n-1} < 0.0001$ or the number of generation reaches the maximum number of generations T_{max}.

Simulation analysis

Performance measures

In this article, the range of CRI is set as [0, 1]. In order to evaluate the accuracy of the model, mean average error (MAE) and root mean square error (RMSE) are adopted, which are represented as

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| \widehat{CRI}(i) - CRI(i) \right| \times 100\% \qquad (25)$$
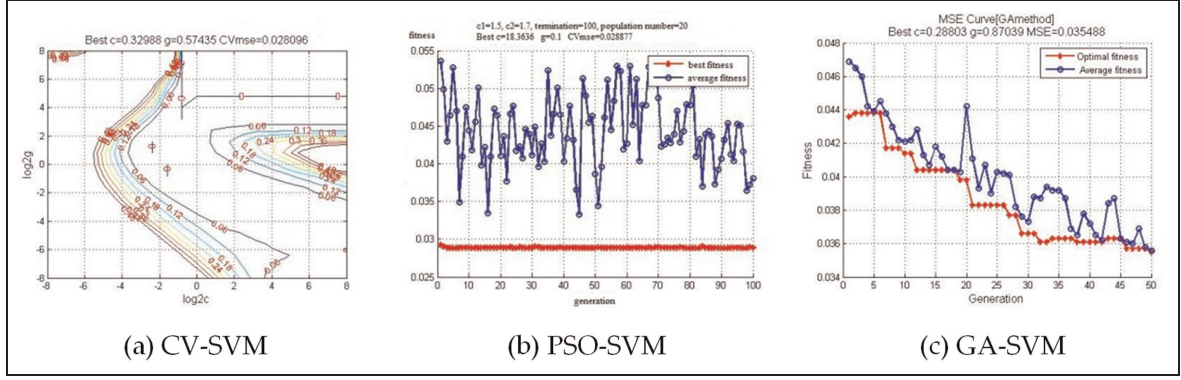
Figure 5. The parameter optimization results of SVM models: (a) CV-SVM, (b) PSO-SVM, and (c) GA-SVM.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(\widehat{CRI}(i) - CRI(i) \right)^2}{n-1}} \qquad (26)$$

where n is the number of testing sample, $\widehat{CRI}(i)$ is the estimation result of the proposed model, and $CRI(i)$ is the actual value.

Data selection

In the case of situation that two vessels encounter under the environment of good visibility, *COLREGS* divides the encounter situation into three situations: head-on, crossing, and overtaking. When vessel is implementing collision avoidance, considering the different features coming from vessels on different bearing, and in order to satisfy the requirements of collision avoidance, the action situation can be divided into six situations (i.e. A, B, C, D, E, and F), just as shown in Figure 2. Considering it is difficult to obtain enough data which can comprehensively reflect all action situations, this article uses simulation data samples for the presented model test. In which, the input feature vector for SVM $V_{in} = [\varphi_0, \varphi_T, V_0, V_t, \theta_T, D_R]$ is obtained with the aid of computer.

In the process of generating sample randomly, in order to reflect the real encounter situation between vessels and guarantee the effectiveness of data set, the range of parameters in actual practice should be considered. According to the collision avoidance situation and the relationship between the relative bearing and the critical safety passing distance, under head-on and crossing conditions, the initial distance between two vessels preferably is 6 nautical miles (nm). So, the range of D_R is set to [0, 7] nautical miles (nm). The velocity of vessel depends on the vessel type. Generally, the velocity of dry bulk vessels and oil tankers is slower, ranging from 13 to 17 knots, and the container vessels faster. The highest velocity can reach 20–23 Kn. But in fact, container vessels generally adapt economical speed

(i.e. 18 Kn). Comprehensively considering the type of vessels and navigation situation, the range of V_0, V_T are set to [6, 20] nm/h and ϕ_0, ϕ_T, and θ_T are set to [0°, 360°].

According to the range discussed above, samples data can be generated. In order to ensure the effectiveness of data set, samples date is collected from six encounter situations (A–F), in which each situation contains 50 groups' data. So, it's totally 300 groups' data. And then we choose 50 groups data as testing sample.

Results analysis

In the test experiment, to properly optimize the three parameters C, γ, and ε for SVM, GA is used. Before the implementation, four GA parameters, namely, p_c, p_m, p_{size} and T_{max}, need to be predetermined. In general, p_c varies from 0.3 to 0.9, p_m varies from 0.01 to 0.1, and p_{size} is the population size which is set according to the size of the samples. T_{max} is the maximum number of generations. For contrast, the parameters optimization process by CV and PSO are compared with GA, as shown in Figure 5. The test results are listed in Table 2 and are shown in Figure 6.
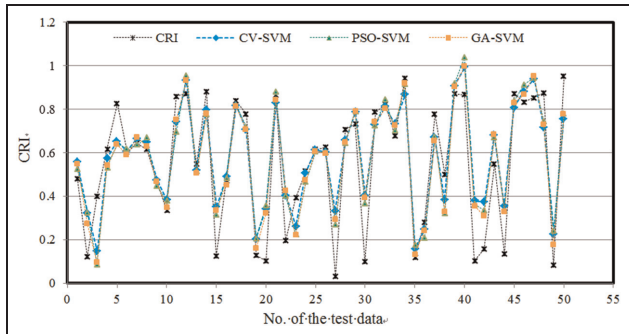
According to comparison results, it is obvious that the prediction accuracy of GA-based method is improved than cross validation (CV) method which is often adopted in SVM training process. Furthermore, for computational speed for parameters searching, the time consumption of CV-SVM model is shortest. However, by comparison with the aviation and road transportation, the velocity and encounter situation of vessels change slowly and the process of collision avoidance needs 1–2 h. So CRI estimation doesn't have strict demand on time consumptions and in actual practice the parameters optimization model with high accuracy should be chosen.

MAE and RMSE demonstrate the error between the estimation value and the real value of CRI. They can be used to evaluate estimation accuracy. But on the

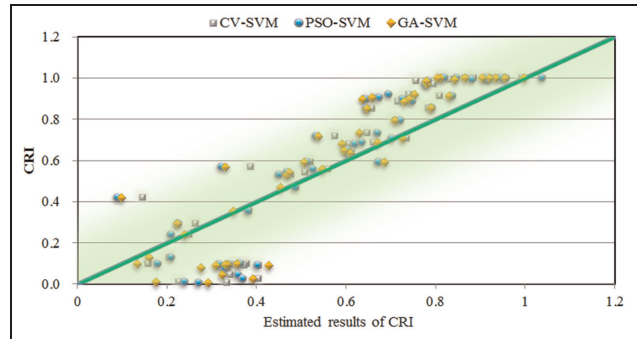**Table 2.** Comparison with different parameter search results.

Method	Best c	Best g	Testing time (ms)	T_{max}	SVTotal	MAE	RMSE
CV-SVM	0.32988	0.57435	0.47	–	61	0.16139	0.19333
PSO-SVM	18.3636	0.1	2.47	100	44	0.14986	0.18009
GA-SVM	0.28803	0.8704	1.22	100	32	0.13744	0.16907

MAE: mean average error; RMSE: root mean square error; CV-SVM: cross-validation support vector machine; PSO-SVM: particle swarm optimization support vector machine; GA-SVM: genetic algorithm support vector machine.
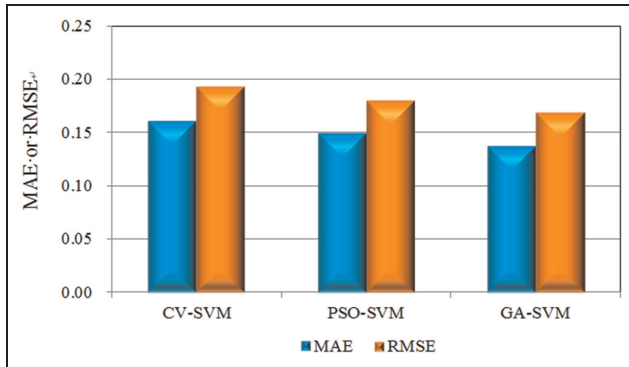
Figure 6. Error comparison of SVM models.

Figure 8. Estimation ability of the three models on CRI.

Figure 7. The estimative result of SVM models.

contrary, they can't indicate the estimative value is less than or more than the real value. The accurate results can directly affect the opportunity of taking collision avoidance plan. So, we give out the estimative value of CRI for 50 groups testing data in Figure 7. And the CRI scatterplot between the estimative value and real value is shown in Figure 8.

From Figure 7, we can see that 50 groups testing data probably cover all the CRI situation between 0 and 1. Figure 8 shows that when CRI is less than 0.43, the estimative value is more than the real value. And when CRI is not less than 0.43, the estimative value is less than the real value. Zhou and Wu[10] divided thevessel collision risk into three levels according to CRI: "low risk"($0 \leq CRI < 0.4$), "moderate

risk"($0.4 \leq CRI < 0.7$), and "high risk"($0.7 \leq CRI \leq 1$). For the estimative results, if vessels' risk is lower, the higher estimative value doesn't have huge influences on navigation and collision avoidance plan. But if the vessels' risk is higher, especially when the value of CRI is not less than 0.7, the lower estimative may influence the risk rate and the opportunity of taking collision avoidance plan. So, the rating threshold can be appropriately adjusted according to actual situation when applying this model to the judgment of vessel collision risk rating.

Conclusion

This article proposes a CRI estimation model based on SVM and applies GAs to optimize the corresponding parameters. And then the verification and analysis with simulation samples data are conducted by comparisons between CA-SVM, PSO-SVM, and GA-SVM models. Results show that the CRI estimation model based on SVM has higher accuracy and the accuracy of GA-SVM model is the best. However, the estimative value of SVM model is lower when CRI is higher. So, the rating threshold can be appropriately adjusted according to actual situation when applying this model to the judgment of vessel collision risk rating.

However, there are some unresolved issues to be discussed in the future work. First of all, this article applies simulation analysis to the training and testing of the model. In the future study, in order to test the accuracy

and applicability of model in practical collision avoidance, actual vessel data should be applied to the model testing procedure.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

References

1. Kearon J. Computer programs for collision avoidance and traffic keeping. In: *Conference on mathematical aspects on marine traffic*. London: Academic Press, 1977.

2. Vahidi A and Eskandarian A. Research advances in intelligent collision avoidance and adaptive cruise control. *IEEE Trans Intell Transp Syst* 2003; 4: 143–153.

3. Lee HJ and Rhee KP. Development of collision avoidance system by using expert system and search algorithm. *Int Shipbuilding Prog* 2001; 48: 197–212.

4. Liu YH and Liu HX. Case learning base on evaluation system for vessel collision avoidance. In: *Proceedings of international conference on machine learning and cybernetics*, Dalian, China, 13–16 August 2006. New York: IEEE.

5. Ren YL, Mou JM, Yan QX, et al. Study on assessing dynamic risk of ship collision. In: *Proceedings of the first international conference on transportation information and safety*, Wuhan, China, 30 June–2 July 2011, pp.2751–2757. New York: IEEE.

6. Ahn JH, Rhee KP and You YJ. A study on the collision avoidance of a ship using neural networks and fuzzy logic. *Appl Ocean Res* 2012; 37: 162–173.

7. Xu Q. Collision avoidance strategy optimization based on danger immune algorithm. *Comput Ind Eng* 2014; 76: 268–279.

8. Cockeroft AN. The circumstance of sea collision. *J Navigation* 1982; 35: 100–112.

9. Kao SL, Lee KT, Chang KY, et al. A fuzzy logic method for collision avoidance in vessel traffic service. *J Navigation* 2007; 60: 17–31.

10. Zhou JH and Wu CJ. Construction of the collision risk factor model. *J Ningbo U* 2004; 17: 61–65.

11. Fei L, Wei N and Song HW. Integrated closeness based on generalized fuzzy and the application in risk analysis. *Adv Syst Sci Appl* 2005; 5: 111–117.

12. Kim SG and Kim YG. An autonomous navigation system for unmanned underwater vehicle. In: Inzartsev AV (ed.) *Underwater vehicle* 2009, pp.279–294. ISBN: 978-953-7619-49-7.

13. Bukhari AC, Tusseyeva I, lee BG, et al. An intelligent real-time multi-vessel collision risk assessment system from VTS view point based on fuzzy inference system. *Expert Syst Appl* 2013; 40: 1220–1230.

14. Chen JH, Chen HW and Liu K. A method of estimating ship collision risk based on fuzzy neural network. *Ship Sci Tech* 2008; 30: 135–138.

15. Su X, Li Z, Feng Y, et al. New global exponentials ability criteria for interval-delayed neural networks. *Proc IMechE, Part I: J Systems and Control Engineering* 2011; 255: 125–136.

16. Zhu X, Xu H and Lin J. Domain and its model based on neural networks. *J Navigation* 2001; 54: 97–103.

17. Shi P, Zhang YQ and Agarwal RK. Stochastic finite time state estimation for discrete time-delay neural networks with Markovian jumps. *Neurocomputing* 2015; 151: 168–174.

18. Shi P, Zhang Y, Chadli M, et al. Mixed H-infinity and passive filtering for discrete fuzzy neural networks with stochastic jumps and time delays. *IEEE Trans Neural Netw Learn Syst* 2015; 27: 903–909.

19. Zhang ZH, Yue SG and Zhang GP. Fly visual system inspired artificial neural network for collision detection. *Neurocomputing* 2015; 153: 221–234.

20. Chin HC and Debnath AK. Modeling perceived collision risk in port water navigation. *Safety Sci* 2009; 47: 1410–1416.

21. Mou JM, van der Tak C and Ligteringen H. Study on collision avoidance in busy water ways by using AIS data. *Ocean Eng* 2010; 37: 483–490.

22. Perera LP, Carvalho JP and Guedes Soares C. Intelligent ocean navigation and fuzzy-Bayesian decision/action formulation. *IEEE J Ocean Eng* 2012; 37: 204–219.

23. Yu B, Song XL, Guan F, et al. k-nearest neighbor model for multiple-time-step prediction of short-term traffic condition. *J Transp Eng: ASCE* 2016; 142: 04016018.

24. Yao BZ, Hu P, Zhang MH, et al. A support vector machine with the tabu search algorithm for freeway incident detection. *Int J Appl Math Comp* 2014; 24: 397–404.

25. Yao BZ, Yao JB, Zhang MH, et al. Improved support vector machine regression in multi-step-ahead prediction for rock displacement surrounding a tunnel. *Sci Iran* 2014; 21: 1309–1316.

26. Yao BZ, Chen C, Cao QD, et al. Short-term traffic speed prediction for an urban corridor. *Comput-Aided Civ Inf*. Epub ahead of print 21 July 2016. DOI: 10.1111/mice.12221.

27. Hou SM and Li YR. Short-term fault prediction based on support vector machines with parameter optimization by evolution strategy. *Expert Syst Appl* 2009; 36: 12383–12391.

28. Hsu CW, Chang CC and Lin CJ. *A practical guide to support vector classification*. Technical report, Freiburg, Germany, 15 July 2003. Taipei City, Taiwan: Department of Computer Science and Information Engineering, National Taiwan University.

29. Yu B, Kong L, Sun Y, et al. A bi-level programming for bus lane network design. *Transport Res C* 2015; 55: 310–327.

30. Yu B, Peng ZX, Wang KM, et al. An optimization method for planning the lines and the operational strategies of waterbuses: the case of Zhoushan city. *Oper Res Quart* 2015; 15: 25–49.

31. Yao BZ, Yu B, Hu P, et al. An improved particle swarm optimization for carton heterogeneous vehicle routing problem with a collection depot. *Ann Oper Res* 2016; 242: 303–320.

32. Konak A, Coit DW and Smith AE. Multi-objective optimization using genetic algorithms: a tutorial. *Reliab Eng Syst Safe* 2006; 91: 992–1007.